

SEQUENCE LISTING

SEIVED OF MAN

OCT 26 2000

TECH GENTER 1608/2900

(1) GENERAL INFORMATION:

(i) APPLICANT: von Schaewen, Antje Dr. rer. nat.

OCT 30 200C

- (ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acet&CHCENTER1600/2990 glucosaminyl transferase I (GnTI) activity
- (iii) NUMBER OF SEQUENCES: 6
  - (iv) CORRESPONDENCE ADDRESS:
    - (A) ADDRESSEE: von Schaewen, Antje Dr. rer. nat.
    - (B) STREET: Natruperstrasse 169a
    - (C) CITY: Osnabrueck
    - (D) COUNTRY: Germany
    - (E) ZIP: D-49076
  - (v) COMPUTER READABLE FORM:
    - (A) MEDIUM TYPE: Floppy disk
    - (B) COMPUTER: IBM PC compatible
    - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
    - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
  - (vi) CURRENT APPLICATION DATA
    - (A) APPLICATION NUMBER: 09/591,466
    - (B) FILING DATE: 06/09/2000
    - (C) CLASSIFICATION: 536
- (2) INFORMATION FOR SEQ ID NO: 1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 1669 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: cDNA to mRNA
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: Solanum tuberosum
    - (B) STRAIN: Desiree
    - (D) DEVELOPMENTAL STAGE: Sink organ
    - (F) TISSUE TYPE: Mesophyll
    - (G) CELL TYPE: Leaf cells
  - (vii) IMMEDIATE SOURCE:
    - (A) LIBRARY: Lambda ZAP II (Eco RI)
    - (B) CLONE: gntI-A1(K)
  - (ix) FEATURE:
    - (A) NAME/KEY: misc\_feature
    - (B) LOCATION: 659..667
    - (D) OTHER INFORMATION:/function= "Asn codon in this context is a potential glycosylation site" /product= "N-glycosylation consensus sequence"

> > 44

45 46

47

48

49

31 32

33

60

61

62

63

```
/phenotype= "N-glycans modulate protein
                       properties"
 2
3
4
                       /standard_name= "N-glycosylation site"
                       /label= pot-CHO
                       /note= "GnTI-coding sequences from animals do not
 5
6
7
                       contain this feature"
 8
          (ix) FEATURE:
                (A) NAME/KEY: CDS
 9
                (B) LOCATION:53..1393
10
                (C) IDENTIFICATION METHOD: experimental
11
                (D) OTHER INFORMATION:/codon_start= 53
12
                       /function= "initiates complex N-glycans on
13
                       secretory glycoproteins"
14
                       /EC number= 2.4.1.101
15
16
                       /product=
                       "beta-1,2-N-acetylglucosaminyltransferase I"
17
                       /evidence= EXPERIMENTAL
18
                       /gene= "cgl"
19
                        /standard_name= "gntI"
20
                        /label= ORF
21
                        /note= "first gntI sequence from potato (unpublished)"
22
23
24
          (ix) FEATURE:
                 (A) NAME/KEY: 5'UTR
25
                (B) LOCATION: 15..52
26
27
          (ix) FEATURE:
28
                 (A) NAME/KEY: 3'UTR
29
                 (B) LOCATION: 1394..1655
30
31
32
          (ix) FEATURE:
                 (A) NAME/KEY: CDS
33
                 (B) LOCATION:80..139
34
                 (D) OTHER INFORMATION: /function= "membrane anchor (amino
35
                        acids 10-29)"
36
                        /product= "hydrophobic amino acid stretch in GnTI"
37
                        /standard name= "membrane anchor of a type II
38
                        Golgi protein"
39
                        /note= "identified by comparison with GnTI sequences
40
                       from animals"
41
42
          (ix) FEATURE:
43
                 (A) NAME/KEY: misc feature
44
45
                 (B) LOCATION:1..14
                 (D) OTHER INFORMATION:/function= "used for cloning the
46
                        cDNA library in Lambda ZAPII"
47
                        /product= "EcoRI/NotI-cDNA adapter"
48
                        /number= 1
49
50
 51
           (ix) FEATURE:
                 (A) NAME/KEY: misc feature
 52
                 (B) LOCATION: 1656...1669
 53
                 (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
 54
                         /number= 2
 55
 56
 57
           (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 58
 59
      GAATTCGCGG CCGCCTGAGA AACCCTCGAA TTCAATTTCG CATTTGGCAG AG ATG
 60
                                                                    Met
 61
 62
```

63

55

•					•				:	3									
1 AGA 2 Arg 3	GG G1	GG A .y A	AC A	AAG Lys 5	TTT Phe	TGC ' Cys	TTT (	GAT Asp	TTA Leu 10	CGG Arg	TAC Tyr	CTT Leu	CTC Leu		C G	STG /al	GCT Ala	C a	103
GCT Ala	C1 L∈	rc G eu A	GCC '		ATC Ile	TAC Tyr	ATA (	CAG Gln 25	ATG Met	CGG Arg	CTT Leu	TTC Phe	GCG Ala	_	CA (	CAG Gln	TCA Sei	A r	151
GAA Glu	T	AT G yr V 35		GAC Asp	CGC Arg	CTT Leu	GCT Ala 40	GCT Ala	GCA Ala	ATT Ile	GAA Glu	GCA Ala 45	0.4.0	A AA	AT ( sn !	CAT	TG' Cy:	T s	199
ACA Thr 50	S	GT ( er (	CAG Gln	ACC Thr	AGA Arg	TTG Leu 55	CTT Leu	ATT Ile	GAC Asp	AAG Lys	ATT Ile 60	501	CA(	G C	AG ln	CAA Gln	GG. G1 6	A Y 5	247
AGA Arç	G J V	TA (	GTA Val	GCT Ala	CTT Leu 70	GAA Glu	GAA Glu	CAA Gln	ATG Met	AAG Lys 75	1110	CAC Glr	GA As	C C p G	AG ln	GAG Glu 80	TG Cy	C 's	295
CG( Ar	G C g G	AA '	TTA Leu	AGG Arg 85	GCT Ala	CTT Leu	GTT Val	CAG Gln	GAT Asp 90	CTT Leu	GAA Glu	AG:	r AA c Ly	G G s G	GC Sly 95	ATA Ile	AA Ly	AA /s	343
AA( Ly:	G T s L	₋eu	ATC Ile 100	GGA Gly	GAT Asp	GTG Val	CAG Gln	ATG Met 105	CCA Pro	GTG Val	GCF Ala	A GC' a Al	r GT a Va 11		TT /al	GTT Val	ra Me	rG ∋t	391
Al	a (	Cys 115	Ser	Arg	Thr	GAC Asp	120	ьeu	GIL	LALC	, 111.	12	5	, ,					439
Lу 13	ន : 0	Гуr	Gln	Thr	: Sei	GTT Val	Ala	Ser	. г.	3 I Å 1	14	0	u 11		120		1	45	487
As	p (	Gly	Ser	Asr	1 Pro		o vai	. Arc	і гу	15	5	a no	.u b	O-1	- 1 -	160	ć		535
Le	eu	Thr	Туг	Met 16	t G1: 5	G CAO	s Lev	ı Ası	17	0	u rr	0 00			175	•		_	583
Pı	0	Gly	Gl: 180	Le <sup>1</sup>	u Va	T GC	a Tyi	18	г шу 5	SIL	e Ai	a A	1	90	-1-	· -1		L	631
A.	la	Leu 195	Ası	p Gl	n Le	G TT u Ph	e H1:	s Ly О	S HI	S AS	) II F1	2	05	9					679 727
L 2	eu 10	Glu	ı As	p As	р Ме	G GA et Gl 21	u 11 5	e Aı	а Ал	.a A	2	20		TO P	- 1		:	225	775
8 A 9	la	Gl	y Al	a Th	nr Le 23	TT CI eu Le 30	eu As	p Ar	g As	2: 2:	35 35	CI I		100		24	40		823
	CT	TG( Tr	G AA p As	n As	AC AZ sp A: 45	AT GO sn Gl	GA CA Ly Gl	A AC	rg G.	AG T' ln P' 50	TC G he V	TC Cal G	CAA (	GAT Asp	CC Pr 25		sp	Ala	023

1 2 3 4	CTT Leu	TAC Tyr	CGC Arg 260	TCA Ser	GAC Asp	TTT Phe	TTT Phe	CCT Pro 265	GGT Gly	CTT Leu	GGA Gly	TGG Trp	ATG Met 270	CTT Leu	TCA Ser	AAA Lys	871
5 6 7 8	TCA Ser	ACT Thr 275	TGG Trp	TCC Ser	GAA Glu	CTA Leu	TCT Ser 280	CCA Pro	AAG Lys	TGG Trp	CCA Pro	AAG Lys 285	GCT Ala	TAC Tyr	TGG Trp	GAT Asp	919
9 10 11 12	GAC Asp 290	TGG Trp	CTA Leu	AGG Arg	CTG Leu	AAA Lys 295	GAA Glu	AAT Asn	CAC His	AGA Arg	GGT Gly 300	CGA Arg	CAA Gln	TTT Phe	ATT Ile	CGC Arg 305	967
13 14 15 16	CCA Pro	GAA Glu	GTT Val	TGC Cys	AGA Arg 310	ACG Thr	TAC Tyr	AAT Asn	TTT Phe	GGT Gly 315	GAG Glu	CAT His	GGT Gly	TCT Ser	AGT Ser 320	TTG Leu	1015
17 18 19 20	GGG Gly	CAG Gln	TTT Phe	TTT Phe 325	AAG Lys	CAG Gln	TAT Tyr	CTT Leu	GAG Glu 330	CCA Pro	ATT Ile	AAG Lys	CTA Leu	AAT Asn 335	GAT Asp	GTC Val	1063
21 22 23 24	CAG Gln	GTT Val	GAT Asp 340	TGG Trp	AAG Lys	TCA Ser	ATG Met	GAC Asp 345	CTA Leu	AGT Ser	TAC Tyr	CTT Leu	TTG Leu 350	GAG Glu	GAC Asp	AAC Asn	1111
25 26 27 28	TAT Tyr	GTG Val 355	AAA Lys	CAC His	TTT Phe	GGC Gly	GAC Asp 360	TTG Leu	GTT Val	AAA Lys	AAG Lys	GCT Ala 365	AAG Lys	CCC Pro	ATC Ile	CAC His	1159
29 30 31 32	GGA Gly 370	GCT Ala	GAT Asp	GCT Ala	GTT Val	TTG Leu 375	AAA Lys	GCA Ala	TTT Phe	AAC Asn	ATA Ile 380	Asp	GGT Gly	GAT Asp	GTG Val	CGT Arg 385	1207
33 34 35 36	ATT Ile	CAG Gln	TAC Tyr	AGA Arg	GAC Asp 390	CAA Gln	CTA Leu	GAC Asp	TTT Phe	GAA Glu 395	Asp	ATC Ile	GCT Ala	CGA Arg	CAG Gln 400	TTT Phe	1255
37 38 39 40	GGC Gly	ATT	TTT Phe	GAA Glu 405	Glu	TGG Trp	AAG Lys	GAT Asp	GGT Gly 410	Val	CCA Pro	CGG Arg	GCA Ala	GCA Ala 415	Tyr	AAA Lys	1303
41 42 43 44	GGG Gly	ATA Ile	GTA Val 420	Val	TTC Phe	CGG Arg	TTT Phe	CAA Glr 425	Thr	TCT Ser	AGA Arç	A CGI J Arg	GTG Val 430	. Phe	CTT Leu	GTT Val	1351
45 46 47 48	TCC Ser	CCT Pro 435	GAT Asp	TCT Ser	CTT	CGA Arg	CAA Glr 440	Leu	GGA Gly	GTT Val	GA <i>F</i> Glu	A GAT 1 Asp 445	Thr	TAC	3		1393
49 50	CGA	AGAI	TATG	ATTG	GAGC	CT G	AGC <i>F</i>	ACA	TT TA	AGAC	CTTAT	TTC	GTAC	GAT	ACAT	TTGAAA	1453
51 52	GAG	CTGF	ACAC	GAAA	AAGTA	TG A	CTAC	CCAGI	ra go	CTAC	ATGC	A ACA	TTTT	TAAT	GTTA	AATGGAA	1513
53 54	GGA	ACC	CACT	GCTT	TATT	TT G	GAAT	rgga:	rg A <i>F</i>	ATCAT	CAC	C ACA	ATCC	TTAT	ATTO	CAAGTTT	1573
55 56	ACA	AAC	AATA	AGA	GGAAA	ATG I	TGC	CCTA!	ra az	AAACA	TAAL	r TT	rTGT:	TTCT	AAGA	AAGGAAC	1633
5 <b>7</b> 58	GTT	ACG	ATTA	TGAG	GCAAC	CTT I	GGC	GCC	GC GA	TTA	C						1669
59 60	(2	1I (S	NFORI	ITAN	ON FO	OR SE	II QE	ои с	: 2:								
61 62 63			(i)	SEQUAL (A)	JENCI LENG:	E CHA	ARAC' 146	reri: amin	STIC:	S: ids							

		(B (D	) TC	PE: POLO	GY:	line	ar							
	(ii) (xi)	MOL SEQ	ECUL	E TY	PE: SCRI	prot PTIC	ein N: S	EQ I	D NC	): 2:				
Met 1	Arg	Gly	Asn	Lys 5	Phe	Cys	Phe	Asp	Leu 10	Arg	Tyr	Leu	Leu	Val 15
Ala	Ala	Leu	Ala 20	Phe	Ile	Tyr	Ile	Gln 25	Met	Arg	Leu	Phe	Ala 30	Thr
Ser	Glu	Tyr 35	Val	Asp	Arg	Leu	Ala 40	Ala	Ala	Ile	Glu	Ala 45	Glu	Asn
Cys	Thr 50	Ser	Gln	Thr	Arg	Leu 55	Leu	Ile	Asp	Lys	Ile 60	Ser	Gln	Gln
Gly 65	Arg	Val	Val	Ala	Leu 70	Glu	Glu	Gln	Met	Lys 75	His	Gln	Asp	Gln
Cys	Arg	Gln	Leu	Arg 85	Ala	Leu	Val	Gln	Asp 90	Leu	Glu	Ser	Lys	Gly 95
Lys	Lys	Leu	Ile 100	Gly	Asp	Val	Gln	Met 105	Pro	Val	Ala	Ala	Val 110	Val
Met	Ala	Cys 115	Ser	Arg	Thr	Asp	Tyr 120	Leu	Glu	Arg	Thr	Ile 125	Lys	Ser
Leu	Lys 130	Tyr	Gln	Thr	Ser	Val 135	Ala	Ser	Lys	Tyr	Pro 140	Leu	Phe	Ile
Gln 145	Asp	Gly	Ser	Asn	Pro 150	Asp	Val	Arg	Lys	Leu 155	Ala	Leu	Ser	Tyr
Gln	Leu	Thr	Tyr	Met 165	Gln	His	Leu	Asp	Tyr 170	Glu	Pro	Val	His	Thr 175
Arg	Pro	Gly	Glu 180		Val	Ala	Tyr	Tyr 185	Lys	Ile	Ala	Arg	His 190	Туг
Trp	Ala	Leu 195		Gln	Leu	Phe	His 200	Lys	His	Asn	Phe	Ser 205	Arg	Val
Ile	Leu 210		Asp	Asp	Met	Glu 215	Ile	Ala	Ala	Asp	Phe 220	Phe	Asp	Туг
Glu 225		Gly	Ala	Thr	Leu 230		Asp	Arg	Asp	Lys 235	Ser	Ile	Met	Ala
Ser	Ser	Trp	Asn	Asp 245		Gly	Gln	Arg	Gln 250	Phe	Val	Gln	Asp	255
Ala	Leu	Tyr	Arg 260		Asp	Phe	Phe	Pro 265	Gly	Leu	Gly	Trp	270	Let
Lys	s Ser	Thr 275		Ser	Glu	Leu	Ser 280		Lys	Trp	Prc	285	a Ala	а Ту:
Asp	290		Leu	a Arg	, Leu	Lys 295		a Asn	n His	arg	300 300	Arg	g Glr	n Phe

·

```
Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
 2
 3
     Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
 4
                                           330
 5
                      325
 6
     Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
 7
                                       345
 8
 9
     Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
10
                                   360
11
              355
12
     His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
13
14
15
     Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln
16
                           390
                                                395
17
18
     Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
19
20
                      405
21
     Lys Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu
22
                                       425
23
24
     Val Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr
25
26
                                   440
27
28
      (2) INFORMATION FOR SEQ ID NO: 3:
29
30
           (i) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 1737 base pairs
31
                (B) TYPE: nucleic acid
32
                (C) STRANDEDNESS: double
33
34
                 (D) TOPOLOGY: linear
35
          (ii) MOLECULE TYPE: cDNA to mRNA
36
37
38
         (iii) HYPOTHETICAL: NO
39
40
          (iv) ANTI-SENSE: NO
41
42
          (vi) ORIGINAL SOURCE:
                 (A) ORGANISM: Nicotiana tabacum
43
                 (B) STRAIN: Samsun NN
44
45
                 (D) DEVELOPMENTAL STAGE: Sink organ
                 (F) TISSUE TYPE: Mesophyll
46
                 (G) CELL TYPE: Leaf cells
47
48
49
         (vii) IMMEDIATE SOURCE:
                 (A) LIBRARY: Lambda ZAP II (Eco RI)
50
                 (B) CLONE: gntI-A9(T)
51
52
53
          (ix) FEATURE:
                 (A) NAME/KEY: misc_feature
(B) LOCATION:733..741
54
55
                 (D) OTHER INFORMATION:/function= "Asn codon in this
56
                        context is a potential glycosylation site"
57
                        /product= "N-glycosylation consensus sequence"
58
                        /phenotype= "N-glycans modulate protein
59
                        properties"
60
                        /standard name= "N-glycosylation site"
61
                        /label= pot-CHO
62
                        /note= "GnTI sequences from animals do not contain this
63
```

```
feature"
 2
         (ix) FEATURE:
 3
                (A) NAME/KEY: CDS
4567
                (B) LOCATION: 127..1467
                (C) IDENTIFICATION METHOD: experimental
                (D) OTHER INFORMATION:/codon_start= 127
                       /function= "initiates complex N-glycans on
 8
                       secretory glycoproteins"
 9
                        /EC_number= 2.4.1.101
10
                        /product=
11
                        "beta-1,2-N-acetylglucosaminyltransferase I"
12
                        /evidence= EXPERIMENTAL
13
                        /gene= "cgl"
14
                        /standard name= "gntI"
15
                        /label = ORF
16
                        /note= "first gntI sequence from tobacco (unpublished)"
17
18
          (ix) FEATURE:
19
                 (A) NAME/KEY: 5'UTR
20
                 (B) LOCATION: 15..126
21
22
          (ix) FEATURE:
23
                 (A) NAME/KEY: 3'UTR
24
                 (B) LOCATION: 1468...1723
25
26
          (ix) FEATURE:
27
                 (A) NAME/KEY: CDS
28
                 (B) LOCATION: 154..213
29
                 (D) OTHER INFORMATION: /function= "membrane anchor (amino
30
                        acids 10-29)"
31
                        /product= "hydrophobic amino acid stretch in GnTI"
32
                        /standard name= "membrane anchor of a type II
33
                        golgi protein"
34
35
           (ix) FEATURE:
36
                 (A) NAME/KEY: misc_feature
37
                 (B) LOCATION: 1..14
38
                 (D) OTHER INFORMATION:/function= "use for cloning the
39
                        cDNA library in Lambda ZAPII"
/product= "EcoRI/NotI-cDNA adapter"
40
41
                         /number= 1
42
43
           (ix) FEATURE:
 44
                 (A) NAME/KEY: misc feature
 45
                 (B) LOCATION: 1724...1737
 46
                 (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
 47
                         /number= 2
 48
 49
 50
           (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 51
 52
      GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CAAAGTAAAT CCAGCGATGA
                                                                                   60
 53
 54
      AACACTCATA ACTGAACACT GAGAGACTAT TCGCTTTCTC CTAAAGCCTT CAATCGAATT
                                                                                  120
 55
 56
       CGCACG ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC
                                                                                  168
 57
              Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu
                       450
 59
 60
       ATC TTG GCT GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG
                                                                                  216
 61
       Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala
 62
                                         470
 63
                    465
```

1 2 3 4	ACA Thr	CAG Gln.	TCA Ser 480	GAA Glu	TAT Tyr	GCA Ala	GAT Asp	CGC Arg 485	CTT Leu	GCT Ala	GCT Ala	GCA Ala	ATT Ile 490	GAA Glu	GCA Ala	GAA Glu	264	
5 6 7 8	AAT Asn	CAT His 495	TGT Cys	ACA Thr	AGC Ser	CAG Gln	ACC Thr 500	AGA Arg	TTG Leu	CTT Leu	ATT Ile	GAC Asp 505	CAG Gln	ATT Ile	AGC Ser	CTG Leu	312	
9 10 11 12	CAG Gln 510	CAA Gln	GGA Gly	AGA Arg	ATA Ile	GTT Val 515	GCT Ala	CTT Leu	GAA Glu	GAA Glu	CAA Gln 520	ATG Met	AAG Lys	CGT Arg	CAG Gln	GAC Asp 525	360	
13 14 15 16	CAG Gln	GAG Glu	TGC Cys	CGA Arg	CAA Gln 530	TTA Leu	AGG Arg	GCT Ala	CTT Leu	GTT Val 535	CAG Gln	GAT Asp	CTT Leu	GAA Glu	AGT Ser 540	AAG Lys	408	<b>:</b>
17 18 19 20	GGC Gly	ATA Ile	AAA Lys	AAG Lys 545	TTG Leu	ATC Ile	GGA Gly	AAT Asn	GTA Val 550	CAG Gln	ATG Met	CCA Pro	GTG Val	GCT Ala 555	GCT Ala	GTA Val	456	5
21 22 23 24	GTT Val	GTT Val	ATG Met 560	GCT Ala	TGC Cys	AAT Asn	CGG Arg	GCT Ala 565	GAT Asp	TAC Tyr	CTG Leu	GAA Glu	AAG Lys 570	ACT Thr	ATT Ile	AAA Lys	504	1
25 26 27 28	TCC Ser	ATC Ile 575	Leu	AAA Lys	TAC Tyr	CAA Gln	ATA Ile 580	TCT Ser	GTT Val	GCG Ala	TCA Ser	AAA Lys 585	Tyr	CCT Pro	CTT Leu	TTC Phe	552	2
29 30 31 32	ATA Ile 590	Ser	CAG Gln	GAT Asp	GGA Gly	TCA Ser 595	CAT His	CCT Pro	GAT Asp	GTC Val	AGG Arg 600	гÀг	CTT Leu	GCT Ala	TTG Leu	AGC Ser 605	600	0
33 34 35 36	TAT Tyr	GAI Asp	CAG Gln	CTG Leu	ACG Thr 610	Tyr	ATG Met	CAG Gln	CAC His	TTG Leu 615	Asp	TTT Phe	GAA Glu	CCT Pro	GTG Val 620	CAT His	648	8
37 38 39 40	ACT Thr	GA <i>F</i> Glu	AGA Arg	CCA Pro 625	Gly	GAG Glu	CTG Leu	ATT Ile	GCA Ala 630	ı Tyr	TAC Tyr	C AAA C Lys	A ATT	GCA Ala 635	HIG	CAT His	69	6
41 42 43 44	TAC Tyr	AAC Lys	G TGG Trp 640	) Ala	TTG Leu	GAT Asp	CAG Gln	CTO Lev 645	ı Ph∈	TAC Tyr	AAG Lys	G CAT	AAT AST 650	PHE	AGC Sei	C CGT Arg	74	4
45 46 47 48	GTT Val	ATO 116 65!	= Ile	A CTA	GAA Glu	GAT Asp	GAT Asp 660	Met	GA <i>I</i> Glu	A ATT	GCC Ala	C CCT a Pro 665	o Asp	TTT Phe	TTT Phe	r GAC e Asp	79	2
49 50 51 52	TT1 Phe 670	e Ph	r GA( e Glu	G GCT ı Ala	GGF Gly	A GCT Ala 675	Thi	CTI Lev	r CT: 1 Let	GA( Asp	C AGA Arg 680	g Ası	C AAC p Lys	G TCC s Se:	AT'	r ATG e Met 685	84	. 0
53 54 55 56	GC1 Ala	r AT a Il	T TC' e Se:	r TC: r Sei	TG( Tr; 69(	) Asr	GAG n Asp	C AA'	r GG n Gl	A CAZ y Gli 69	n ме	G CA	G TT n Phe	r GT e Va	C CA 1 Gl 70	A GAT n Asp 0	88	38
57 58 59	CC: Pro	T TA	T GC' r Al	T CT a Le	и Ту:	C CG( r Ar	TCZ g Se:	A GA' r As	TTT p Ph 71	e Ph	r CC e Pr	C GG o Gl	T CT' y Le	T GG u Gl 71	А тт	G ATG p Met	93	36
61 62 63	CT' Le	T TC u Se	A AA r Ly	A TC' s Se	T AC'	T TG r Tr	G GA	C GA p Gl	A TT u Le	A TC u Se	T CC r Pr	A AA o Ly	G TG	G CC p Pr	A AA o Ly	G GCT s Ala	98	34

1	TAC TGG GAC GAC TGG CTA NAM CAN AND AND THE ARTY TRY ASP ASP TRY Deu Arg Leu Lys Glu Ash His Arg Gly Arg Gln 745  TTT ATT CGC CCA GAA GTT TGC AGA ACA TAT AAT TTT GGT GAG CAT GGT Phe File Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly 750  TCT AGT TTG GGG CAG TTT TTC AAG CAG TAT CTT GAG CCA ATT AAA CTA Ser Ser Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu 770  AAT GAT GTC CAG GTT GAT TGG AGA GTCA ATG GAC CTT AGT TAC CTT TTG ASN ASP Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu 785  GAG GAC AAT TAC GTG AAA CAC TTT GGT GAC TTG GTT AAA AAG GCT AAG Glu Asp Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys 800  CCC ATC CAT GGA GCT GAT GCT GTC TTG AAA GCA TTT AAC ATA GAT GGT Pro Ile His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly 815  GAT GTG CGT ATT CAG TAC AGA GAT CAA CTA GAC TTT GAA AAT ATC GCA ASP Val Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala 830  CGG CAA TTT GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGT GCA ARG GLI TTT GGA AAT ATC GCA ASP Val Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala 830  CGG CAA TTT GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGT GCA ARG GLI Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala 850  GCA TAT AAA GGA ATA GTA GTA TTC CGG TAC CAA ACG TCC AGA CGT GTA ALA TYR Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val 865  TTC CTT GTT GGC CAT GAT TCG CTT CAA CAA CTC GAA ACT TAT GAA GAT ACT Phe Leu Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr 880  TAA CAAAGATATG ATTGCAGGAG CCCGGGCAAA ATTTTGACT TATTGGGTAG  GATGGACACA ACCTATTATT TTAAGTGTTC AGAACATAAA GAGGAAATGT AGCCCTGTAA  AGACTATACA ACCTATTATT TTAAGTGTTC AGAACATAAA GAGGAAATGT AGCCCTGTAA  AGACTATACAC ACCTATTATT TTAAGTGTTC AGAACATAAA GAGGAAATGT AGCCCTGTAA  AGACTATACA ACCTATTATT TTAAGTGTTC AGAACA																
2 3 4 5	TAC Tyr	Trp	GAC Asp	GAC Asp	TGG Trp	CTA Leu	Arg	CTC Leu	AAA Lys	GAG Glu	AAT Asn	HIS	AGA Arg	GGT Gly	CGA Arg	CAA Gln	1032
6 7 8 9	Phe	ATT Ile	CGC Arg	CCA Pro	GAA Glu	Val	TGC Cys	AGA Arg	ACA Thr	TAT Tyr	Asn	TTT Phe	GGT Gly	GAG Glu	CAT His	GIY	1080
10 11 12 13	TCT Ser	AGT Ser	TTG Leu	GGG Gly	Gln	TTT Phe	TTC Phe	AAG Lys	CAG Gln	Tyr	CTT Leu	GAG Glu	CCA Pro	ATT Ile	гуѕ	CTA Leu	1128
14 15 16 17	AAT Asn	GAT Asp	GTC Val	Gln	GTT Val	GAT Asp	TGG Trp	AAG Lys	Ser	ATG Met	GAC Asp	CTT Leu	AGT Ser	TAT	CTT Leu	TTG Leu	1176
18 19 20 21	GAG Glu	GAC Asp	Asn	TAC Tyr	GTG Val	AAA Lys	CAC His	Phe	GGT Gly	GAC Asp	TTG Leu	GTT Val	тйг	AAG Lys	GCT Ala	AAG Lys	1224
22 23 24 25	CCC Pro	Ile	His	GGA Gly	GCT Ala	GAT Asp	Ala	GTC Val	TTG Leu	AAA Lys	GCA Ala	Pne	Asn	ATA Ile	GAT Asp	GGT Gly	1272
26 27 28 29	Asp	GTG Val	CGT Arg	ATT Ile	CAG Gln	Tyr	Arg	GAT Asp	CAA Gln	CTA Leu	Asp	Phe	GAA Glu	AAT Asn	ATC Ile	Ala	1320
30 31 32 33	CGG Arg	CAA Gln	TTT Phe	GGC Gly	Ile	Phe	GAA Glu	GAA Glu	TGG Trp	ь гар	Asp	GGT Gly	GTA Val	CCA Pro	ALG	ALG	1368
34 35 36 37	GCA Ala	TAI Tyr	AAA Lys	: Gly	Ile	GTA Val	GTT Val	TTC Phe	: Arc	l LÀi	CAA Gln	ACC n Thi	TCC Ser	. Arc	, ALC	GTA Val	1416
38 39 40 41	TTC Phe	CTI Lev	ı Val	. Gly	CAT	GAT Asp	TCG Ser	: Lei	ı Glr	A CAZ n Glr	A CTC	C GGA 1 Gly	∤ TT€	e GIU	A GAT a Asp	ACT Thr	1464
42 43 44 45		CA/	\AGAT	ГАТG	ATTO	GCAG	GAG (	CCG	GCA/	AA A'	TTTT	rgac'	г та:	rtgg(	GTAG		1517
46 47	GAT	GCA'	rcga	GCT	GACA	CTA A	AACC	ATGA'	rt t	racci	AGTT	A CA	TACA	ACGT	TTT	AATGTTA	1577
48 49	OAT	CGGA	GGAG	CTC	ACTG:	rtc '	ragt(	GTTG	AA G	GGAT.	ATCG	G CT	TCTT	AGTA	TTG	GATGAAT	1637
50 51	CAT	гсаа	CACA	ACC'	TATTA	TTP	AATT	GTGT'	TC A	GAAC.	ATAA	A GA	GGAA	ATGT	AGC	CCTGTAA	1697
52 53 54	AG?	ACTA	TACA	TGG	GACC	ATC .	ATAA'	TCGC	GG C	CGCG	AATT	С					1737
55 56	(2	) IN	FORM	ATIO	N FO	R SE	Q ID	NO:	4:								
57 58 59 60 61			(i)	(A) (B)	LENG TYPE	TH: : am	ARAC 446 ino : li	amin acid	o ac	S: ids							
62 63		<b>ć</b> )	i) M	OLEC	ULE	TYPE	: pr	otei	n.								

Met 1	Arg	Gly	Asn	Lys 5	Phe	Cys	Cys	Asp	Phe 10	Arg	Tyr	Leu	Leu	Ile 15
	Ala	Val	Ala 20	Phe	Ile	Tyr	Thr	Gln 25	Met	Arg	Leu	Phe	Ala 30	Thr
Ser	Glu	Tyr 35	Ala	Asp	Arg	Leu	Ala 40	Ala	Ala	Ile	Glu	Ala 45	Glu	Asn
Cys	Thr 50	Ser	Gln	Thr	Arg	Leu 55	Leu	Ile	Asp	Gln	Ile 60	Ser	Leu	Gln
Gly 65	Arg	Ile	Val	Ala	Leu 70	Glu	Glu	Gln	Met	Lys 75	Arg	Gln	Asp	Gln
Cys	Arg	Gln	Leu	Arg 85	Ala	Leu	Val	Gln	Asp 90	Leu	Glu	Ser	Lys	Gly 95
Lys	Lys	Leu	Ile 100	Gly	Asn	Val	Gln	Met 105	Pro	Val	Ala	Ala	Val 110	Val
Met	Ala	Cys 115	Asn	Arg	Ala	Asp	Tyr 120	Leu	Glu	Lys	Thr	Ile 125	Lys	Ser
Leu	Lys 130		Gln	Ile	Ser	Val 135	Ala	Ser	Lys	Tyr	Pro 140	Leu	Phe	Ile
Gln 145		Gly	Ser	His	Pro 150	Asp	Val	Arg	Lys	Leu 155	Ala	Leu	Ser	Туг
Gln	Leu	Thr	Tyr	Met 165	Gln	His	Leu	Asp	Phe 170	Glu	Pro	Val	His	Thr 175
Arg	Pro	Gly	Glu 180	Leu	Ile	Ala	Tyr	Tyr 185	Lys	Ile	Ala	Arg	His 190	Ту
Trp	Ala	Leu 195		Gln	Leu	Phe	туr 200	Lys	His	Asn	Phe	Ser 205	Arg	Va.
	210	)		Asp		215	•				220			
225	5			Thr	230					235				
				Asp 245					250	)				23
			260					265	)				270	,
		275	5	o Asp			280	)				203	)	
	29	0		ı Arç		29	5				300	)		
30	5			l Cys	310	)				313	0			
Le	u Gl	y Gl:	n Ph	e Phe	e Ly:	s Gl	n Ty	r Le	u Gl	u Pro	o Ile	e Ly	s Le	ג As

-

```
335
                                           330
                      325
 2
     Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
3
                                       345
 4
                  340
5
     Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
 6
7
 8
     His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
 9
10
11
     Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln
12
13
14
     Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
15
16
                      405
17
     Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu
18
                                       425
19
20
     Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr
21
22
              435
23
24
     (2) INFORMATION FOR SEQ ID NO: 5:
25
26
           (i) SEQUENCE CHARACTERISTICS:
27
                (A) LENGTH: 1854 base pairs
28
                (B) TYPE: Nucleotide
29
30
                (C) STRANDEDNESS: double
                (D) TOPOLOGY: linear
31
32
          (ii) MOLECULE TYPE: cDNA to mRNA
33
34
35
         (iii) HYPOTHETICAL: No
36
          (iv) ANTI-SENSE: No
37
38
          (vi) ORIGINAL SOURCE:
39
                (A) ORGANISM: Arabidopsis thaliana
40
                (B) STRAIN: Columbia
41
                (D) DEVELOPMENTAL STAGE: Mature plants
42
                (F) TISSUE TYPE: All tissues
43
44
45
         (vii) IMMEDIATE SOURCE:
                (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XhoI) and
46
                         Lambda ACT (XhoI)
47
                (B) CLONE: pBSK-Ara-GntI-full #8
48
49
50
          (ix) FEATURE:
                 (A) NAME/KEY: misc_feature
51
                 (B) LOCATION: 1185...1193
52
                 (D) OTHER INFORMATION:/function= "Asn Codon is a
53
                        potential glycosylation site"
54
                        /product= "Consensus sequence for
55
56
                        N-glycosylation"
                        /phenotype= "N glycans modulate
57
                         protein characteristics"
58
                        /standard name= "N glycosylation site"
59
                        /label= pot-CHO
60
                        /note= "absent in animal GnTI sequences"
61
          (ix) FEATURE:
```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	<pre>(A) NAME/KEY: CDS (B) LOCATION:1351469 (C) IDENTIFICATION METHOD: experimental (D) OTHER INFORMATION:/codon_start= 135</pre>	
16 17 18 19	(ix) FEATURE: (A) NAME/KEY: 5'UTR (B) LOCATION:19134	
20 21 22 23	<pre>(ix) FEATURE:     (A) NAME/KEY: 3'UTR     (B) LOCATION:14701848</pre>	
24 25 26 27 28 30 31 33 33 34 56	<pre>(ix) FEATURE:     (A) NAME/KEY: CDS     (B) LOCATION:157215     (D) OTHER INFORMATION:/function= "membrane anchor</pre>	
36 37 38 39 40 41 42 43	<pre>(ix) FEATURE:     (A) NAME/KEY: misc_feature     (B) LOCATION:118     (D) OTHER INFORMATION:/function= "for preparation</pre>	
44 45 46 47 48 49	<pre>(ix) FEATURE:     (A) NAME/KEY: misc_feature     (B) LOCATION:18491854     (D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"     /number= 2</pre>	
51 52 53	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:	
54 55	CTCGAGGCCA CGAAGGCCAC CGTTTTTGTT ATAACGAACG ACACCGTTTC AAACAACTTC	60
56	CTTATTAGCT AGCTCCCTCC CGGCGGCAAA CACCAGAAGA TCCACCGCTT TTGATCTGGT	120
57 58 59 60	TGTTTGTCGT CGAT ATG GCG AGG ATC TCG TGT GAC TTG AGA TTT CTT CTC  Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu  1 5 10	170
61 62 63	ATC CCG GCA GCT TTC ATG TTC ATC TAC ATC CAG ATG AGG CTT TTC CAG Ile Pro Ala Ala Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln	218

1			15					20					25				
2 3 4 5	ACG Thr	CAA Gln 30	TCA Ser	CAG Gln	TAT Tyr	GCA   Ala .	GAT Asp 35	CGC Arg	CTC Leu	AGT Ser	TCC Ser	GCT Ala 40	ATC Ile	GAA Glu	TCT Ser	GAG Glu	266
6 7 8 9	AAC Asn 45	CAT His	TGC Cys	ACT Thr	AGT Ser	CAA Gln 50	ATG Met	CGA Arg	GGC Gly	CTC Leu	ATA Ile 55	GAT Asp	GAA Glu	GTT Val	AGC Ser	ATC Ile 60	314
10 11 12 13	AAA Lys	CAG Gln	TCG Ser	CGG Arg	ATT Ile 65	GTT Val	GCC Ala	CTC Leu	GAA Glu	GAT Asp 70	ATG Met	AAG Lys	AAC Asn	CGC Arg	CAG Gln 75	GAC Asp	362
14 15 16 17	GAA Glu	GAA Glu	CTT Leu	GTG Val 80	CAG Gln	CTT Leu	AAG Lys	GAT Asp	CTA Leu 85	ATC Ile	CAG Gln	ACG Thr	TTT Phe	GAA Glu 90	AAA Lys	AAA Lys	410
18 19 20 21	GGA Gly	ATA Ile	GCA Ala 95	AAA Lys	CTC Leu	ACT Thr	CAA Gln	GGT Gly 100	GGA Gly	CAG Gln	ATG Met	CCT Pro	GTG Val 105	GCT Ala	GCT Ala	GTA Val	458
22 23 24 25	GTG Val	GTT Val 110	ATG Met	GCC Ala	TGC Cys	AGT Ser	CGT Arg 115	GCA Ala	GAC Asp	TAT Tyr	CTT Leu	GAA Glu 120	AGG Arg	ACT Thr	GTT Val	AAA Lys	506
26 27 28 29	TCA Ser 125	GTT Val	TTA Leu	ACA Thr	TAT Tyr	CAA Gln 130	ACT Thr	CCC Pro	GTT Val	GCT Ala	TCA Ser 135	AAA Lys	TAT Tyr	CCT Pro	CTA Leu	TTT Phe 140	554
30 31 32 33	ATA Ile	TCT Ser	CAG Gln	GAT Asp	GGA Gly 145	TCT Ser	GAT Asp	CAA Gln	GCT Ala	GTC Val 150	Ьys	AGC Ser	AAG Lys	TCA Ser	TTG Leu 155	AGC Ser	602
34 35 36 37	TAT Tyr	AAT Asn	CAA Gln	TTA Leu 160	ACA Thr	TAT Tyr	ATG Met	CAG Gln	CAC His 165	TTG Leu	GAT Asp	TTT Phe	GAA Glu	CCA Pro 170	vaı	GTC Val	650
38 39 40 41	ACT Thr	GAA Glu	AGG Arg 175	Pro	GGT Gly	GAA Glu	CTG Leu	ACT Thr 180	Ala	TAC Tyr	TAC Tyr	AAG Lys	ATT Ile 185	: Ата	. CGT Arg	CAC His	698
42 43 44 45	TAC Tyr	AAG Lys 190	Trp	GCA Ala	CTG Leu	GAC Asp	CAG Gln 195	Leu	TTT Phe	Tyr	: AAA : Lys	His	Груг	Phe	AGT Ser	CGA Arg	746
46 47 48 49	GTG Val 205	Ile	ATA	CTA Leu	GAA Glu	GAC Asp 210	Asp	ATC Met	GAA Glu	ATI Ile	GCT Ala 215	Pro	GAC Asp	TTC Phe	TTT Phe	GAT Asp 220	794
50 51 52 53	TAC Tyr	TTI Phe	GAC Glu	G GCT 1 Ala	GCA Ala 225	Ala	AGT Ser	CTC	ATG Met	G GAT : Asp 230	o Arc	G GAT g Asp	AAA Lys	A ACC	235	ATG Met	842
54 55 56 57	GCT Ala	GC A Ala	r TC <i>l</i> a Sei	A TCF Ser 240	rrr	TAA G Asn	' GAT Asp	AAT Asr	GGA n Gly 245	7 GII	G AAC	G CAC	TT' n Ph	T GT0 e Val 250	г пт:	GAT S Asp	890
58 59 60 61	CCC Pro	C TA'	r GCG r Ala 25	a Lei	A ТА( 1 Туі	C CGA	A TCF g Sei	A GAS ASP 260	p Phe	r TT' ≥ Ph	r CC' e Pro	r GG( o Gly	C CT y Le 26	u GI	G TGO y Tr	G ATG o Met	938
62 63	CTC	C AA	G AG	A TC	G AC	r TGC	G GAT	r GA	G TT	A TC	A CC	A AA	G TG	G CC	A AA	G GCT	986

1 2	Leu	Lys 270	Arg	Ser	Thr	Trp	Asp 275	Glu	Leu	Ser	Pro	Lys 280	Trp	Pro	Lys	Ala	
3 4 5 6	TAC Tyr 285	TGG Trp	GAT Asp	GAT Asp	TGG Trp	CTG Leu 290	AGA Arg	CTA Leu	AAG Lys	GAA Glu	AAC Asn 295	CAT His	AAA Lys	GGC Gly	CGC Arg	CAA Gln 300	1034
7 8 9 10	TTC Phe	ATT Ile	GCA Ala	CCG Pro	GAA Glu 305	GTC Val	TGT Cys	AGA Arg	ACA Thr	TAC Tyr 310	AAT Asn	TTT Phe	GGT Gly	GAA Glu	CAT His 315	GGG Gly	1082
11 12 13 14	TCT Ser	AGT Ser	TTG Leu	GGA Gly 320	CAG Gln	TTT Phe	TTC Phe	AGT Ser	CAG Gln 325	TAT Tyr	CTG Leu	GAA Glu	CCT Pro	ATA Ile 330	AAG Lys	CTA Leu	1130
15 16 17 18	AAC Asn	GAT Asp	GTG Val 335	ACG Thr	GTT Val	GAC Asp	TGG Trp	AAA Lys 340	GCA Ala	AAG Lys	GAC Asp	CTG Leu	GGA Gly 345	TAC Tyr	CTG Leu	ACA Thr	1178
19 20 21 22	GAG Glu	GGA Gly 350	AAC Asn	TAT Tyr	ACC Thr	AAG Lys	TAC Tyr 355	TTT Phe	TCT Ser	GGC Gly	TTA Leu	GTG Val 360	AGA Arg	CAA Gln	GCA Ala	CGA Arg	1226
23 24 25 26	CCA Pro 365	ATT Ile	CAA Gln	GGT Gly	TCT Ser	GAC Asp 370	CTT Leu	GTC Val	TTA Leu	AAG Lys	GCT Ala 375	CAA Gln	AAC Asn	ATA Ile	AAG Lys	GAT Asp 380	1274
27 28 29 30	GAT Asp	GAT Asp	CGT Arg	ATC Ile	CGG Arg 385	TAT Tyr	AAA Lys	GAC Asp	CAA Gln	GTA Val 390	Glu	TTT Phe	GAA Glu	. CGC . Arg	ATT Ile 395	GCA Ala	1322
31 32 33 34	GGG Gly	GAA Glu	TTT Phe	GGT Gly 400	Ile	TTT Phe	GAA Glu	GAA Glu	TGG Trp 405	Lys	GAT Asp	GGT Gly	GTG Val	CCA Pro 410	Arg	ACA Thr	1370
35 36 37 38	GCA Ala	TAT Tyr	AAA Lys 415	Gly	GTA Val	. GTG Val	GTG Val	TTT Phe 420	Arg	ATC	CAG Gln	ACA Thr	ACA Thr 425	Arg	CGI Arg	GTA Val	1418
39 40 41 42	TTC Phe	CTG Leu 430	Val	GGG Gly	CCA Pro	GAT Asp	TCT Ser 435	Val	ATG Met	G CAG	G CTT Leu	GGF Gly 440	, Ile	CGA Arç	AAT JAsr	TCC Ser	1466
43 44	TGA	TGC	AAAA	CAT	ATGA	AAGG	aa A	AGAF	GATT	т то	GACC	GCA'	GCF	AGCCI	CCT		1519
45 46	* 445	i															
47 48	TCI	AGC	AGCT	GTT <i>F</i>	GGTT	GT F	TTGI	TAT	T AT	rgga1	GAGT	TTC	GTAGA	AGCG	GTG	GGTTAA	1579
49 50	CTI	TAAC	CAGC	AAGO	SAAGO	CTC T	GGT	GACCA	AG GO	CTGAT	TGGC	C TTA	AGAA(	STTA	TGG	GAACCCC	1639
51 52	TTG	SAAAC	GGT	CAG	GTT	AAA	TATA	TTTC	AG T	rgtt:	TAT	r AG	rgat'	TATC	TTG	rgggtaa	1699
53 54	CTT	ATA	CGAA	TGCA	TAA	CAT :	CTA	rgca(	GT T	rttc:	rtcg:	r cc	CACT'	rgtt	TTG	GCTTCTC	1759
55 56	TAT	rTGC	ragt	GTA	CATA	rct (	CTTC	AAAC	AT G	TACT	AAAT	A AT	GCGT	GTTG	CTT	CAAAGAA	1819
57 58	GTA	AACT	TTTA	TTA	AAAA	AAA i	AAAA	AAAA	AC TO	CGAG							1854
59 60 61 62	(2)	) IN	FORM							C .							
63			(1)	SEQ	UENC.	E CH.	HRAC	TULL	DIIC								

## (2) INFORMATION FOR SEQ ID NO: 6:

(A) LENGTH: 444 amino acids (B) TYPE: Amino acid (D) TOPOLOGY: Linear (ii) MOLECULE TYPE: Protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6: Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu Ile Pro Ala Ala Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln Thr Gln Ser Gln 2.5 Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu Asn His Cys Thr Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile Lys Gln Ser Arg Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp Glu Glu Leu Val Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys Gly Ile Ala Lys Leu Thr Gln Gly Gly Gln Met Pro Val Ala Ala Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys Ser Val Leu Thr Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser Tyr Asn Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met Ala Ala Ser Ser Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro 

1 2 3	Glu 305	Val	Cys	Arg	Thr	Tyr 310	Asn	Phe	Gly	Glu	His 315	Gly	Ser	Ser	Leu	Gly 320
4 5 6	Gln	Phe	Phe	Ser	Gln 325	Tyr	Leu	Glu	Pro	Ile 330	Lys	Leu	Asn	Asp	Val 335	Thr
7 8 9	Val	Asp	Trp	Lys 340	Ala	Lys	Asp	Leu	Gly 345	Tyr	Leu	Thr	Glu	Gly 350	Asn	Tyr
10 11 12	Thr	Lys	Tyr 355	Phe	Ser	Gly	Leu	Val 360	Arg	Gln	Ala	Arg	Pro 365	Ile	Gln	Gly
13 14 15	Ser	Asp 370	Leu	Val	Leu	Lys	Ala 375	Gln	Asn	Ile	Lys	Asp 380	Asp	Asp	Arg	Ile
16 17 18	Arg 385	Tyr	Lys	Asp	Gln	Val 390	Glu	Phe	Glu	Arg	Ile 395	Ala	Gly	Glu	Phe	Gly 400
19 20 21	Ile	Phe	Glu	Glu	Trp 405	Lys	Asp	Gly	Val	Pro 410	Arg	Thr	Ala	Tyr	Lys 415	Gly
22 23 24	Val	Val	Val	Phe 420	Arg	Ile	Gln	Thr	Thr 425	Arg	Arg	Val	Phe	Leu 430	Val	Gly
25 26 27 28	Pro	Asp	Ser 435	Val	Met	Gln	Leu	Gly 440	Ile	Arg	Asn	Ser				